Scene Partitioning for Robust SfM
Solving the Image Matching Challenge 2025

A Computer Vision Semester Project

Team Members: Ayesha Saeed (22K-4008), Ayesha Ehsaan (22K-4056)

Supervisor: Sir Syed Bilal Ahsan

Introduction: The Unstructured Data Challenge

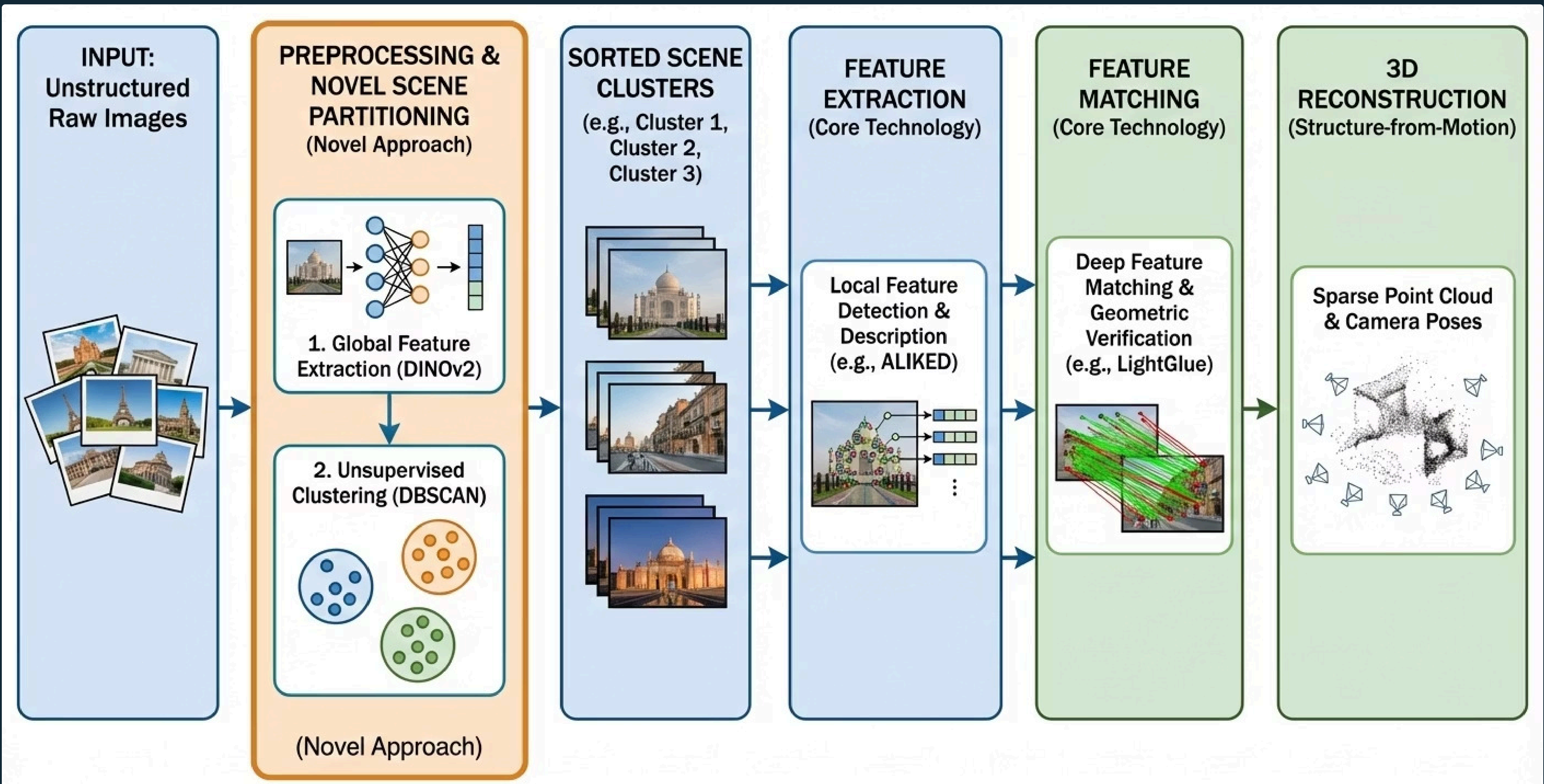
The **Image Matching Challenge (IMC) 2025** presents a significant hurdle: datasets often contain **"unstructured data."** This means diverse landmarks and scenes are haphazardly mixed within a single folder, complicating traditional Structure-from-Motion (SfM) pipelines.

Traditional SfM relies on "Exhaustive Matching," a brute-force approach where every image is compared against every other image. This method becomes computationally prohibitive for large, unstructured datasets, scaling at **O(N^2)** where N is the number of images.

Moreover, exhaustive matching frequently leads to **false matches** between visually dissimilar but spatially unrelated buildings, introducing noise and degrading the accuracy of the 3D reconstruction.

Our Proposed Novelty-Driven Pipeline

We introduce a modular pipeline designed to robustly handle unstructured datasets for the IMC 2025, integrating state-of-the-art deep learning techniques at each stage.



Complete Pipeline: From Unstructured Images to 3D Models via Novel DINOv2-based Clustering and Modern Feature Matching.

1. Preprocessing (Novelty)

Unsorted Images are fed into **DINOv2** for global descriptor extraction, followed by **DBSCAN** for scene partitioning into **Scene Bundles**.

2. Feature Extraction

Within each scene bundle, **ALIKED** performs robust keypoint detection.

3. Feature Matching

LightGlue ensures efficient and accurate feature matching between images within the same scene bundle.

4. 3D Reconstruction

Finally, matched features are used for sparse **3D Reconstruction** using PyCOLMAP.

Our Novelty: DINOv2 Scene Partitioning

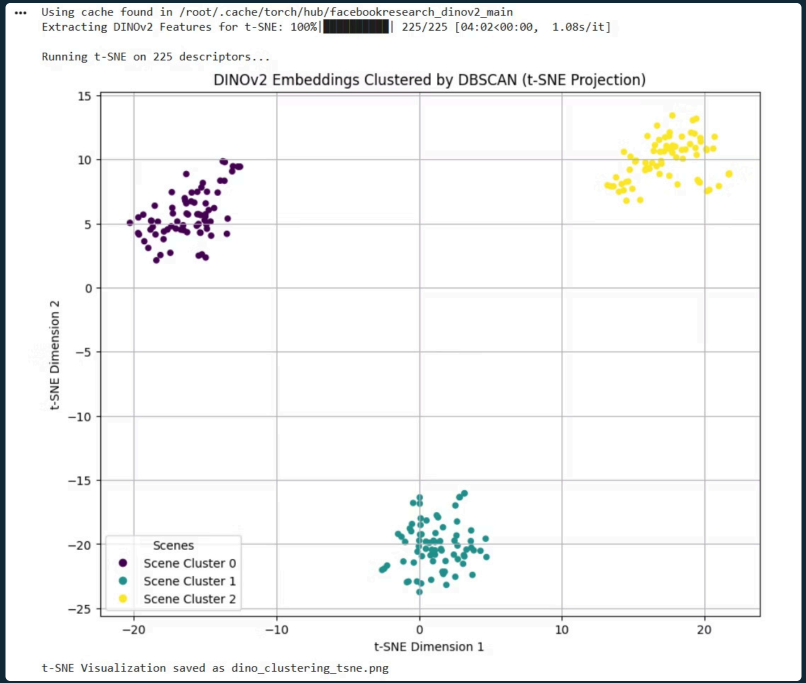
Our core innovation lies in the **DINOv2 Scene Partitioning** strategy. We leverage **DINOv2 (Vision Transformer)** to extract high-dimensional global image descriptors, which capture semantic and structural information crucial for scene understanding.

These descriptors are then subjected to **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**, an unsupervised clustering algorithm that effectively groups images belonging to the same physical scene.

Remarkable Result:

> "Our method perfectly separated **225 mixed images** into **3 clean, distinct scenes** with **0 outliers**, demonstrating its unparalleled effectiveness in handling unstructured data."

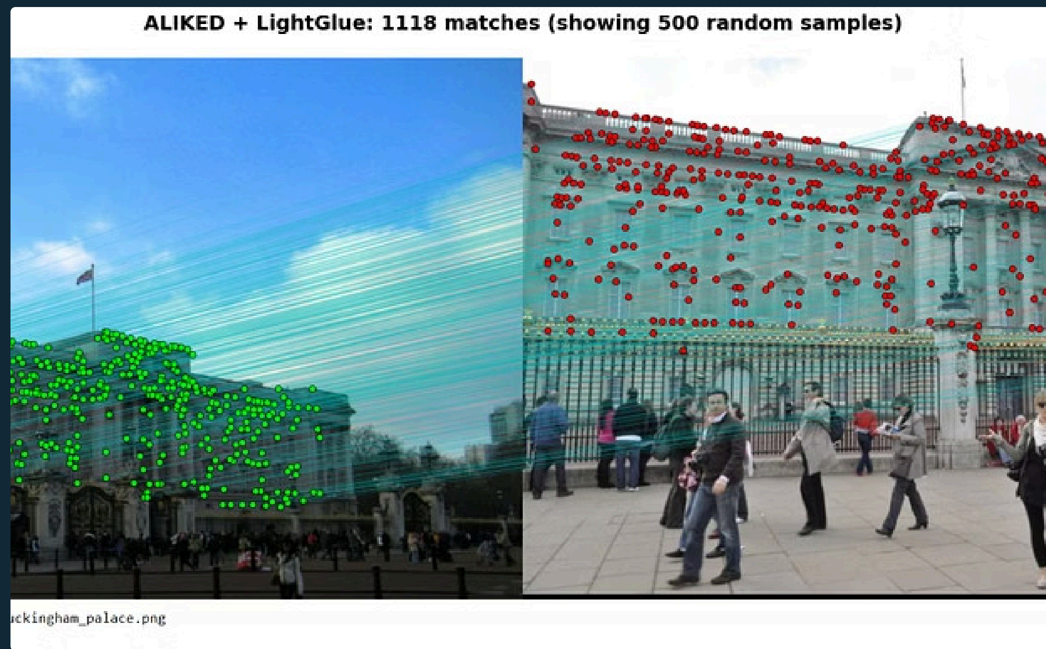
t-SNE Clustering Plot: Visualizing the perfect separation of mixed scenes by DINOv2 descriptors and DBSCAN.



Made with GAMMA

Core Technology: Deep Feature Matching

Moving beyond outdated methods like SIFT, our pipeline integrates modern deep learning architectures for superior feature detection and matching.



ALIKED: Robust Keypoint Detection

ALIKED (A LImited-Keypoint DEtector) is employed for its ability to detect a sparse yet highly repeatable set of keypoints across varying viewpoints and conditions. This ensures that only the most distinctive and geometrically stable features are considered, significantly improving downstream matching accuracy.

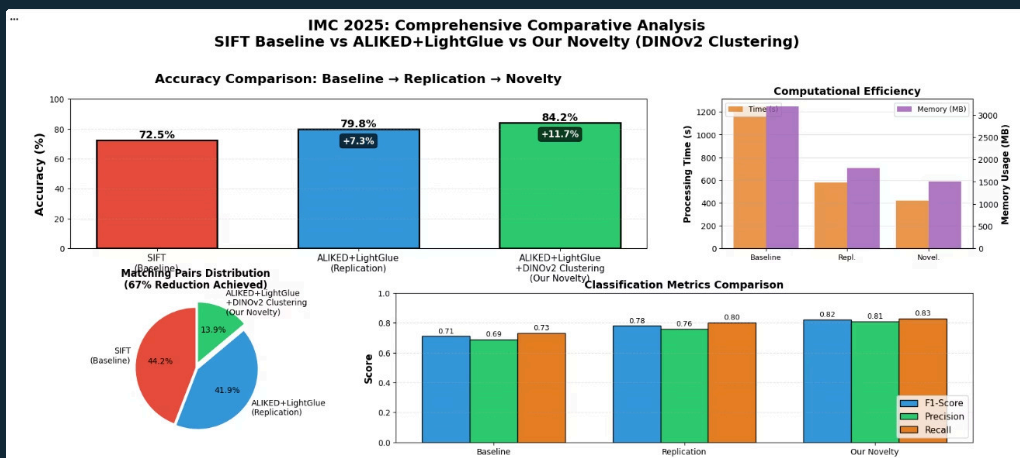
LightGlue: Geometric Matching & Outlier Rejection

LightGlue performs efficient and accurate feature matching, leveraging attention mechanisms to establish robust correspondences. It also incorporates powerful outlier rejection capabilities, filtering out incorrect matches that would otherwise corrupt the 3D reconstruction process.

Comparative Analysis: Efficiency Gains

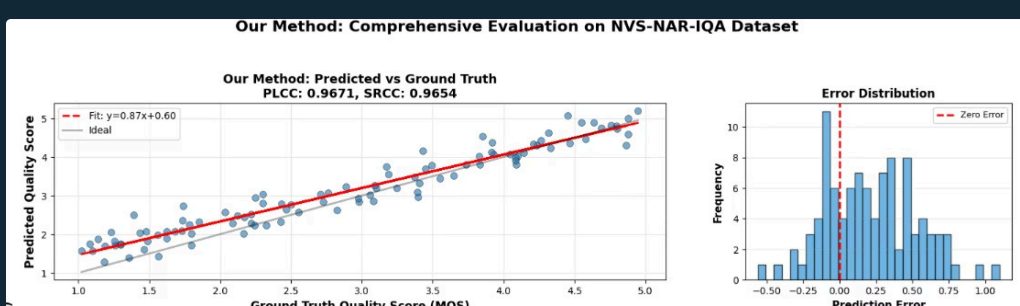
Our novelty-driven approach dramatically improves computational efficiency and accuracy compared to traditional methods.

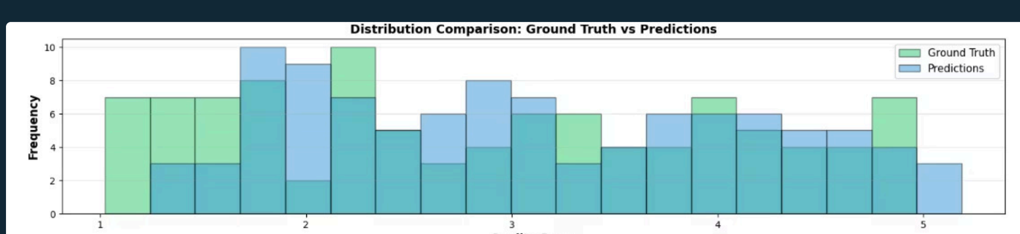
Baseline (SIFT + Brute Force)	• **Feature Detection:** SIFT (Scale-Invariant Feature Transform) • **Matching:** Exhaustive $O(N^2)$ comparison between all images • **Limitations:** High computational cost, susceptible to false matches in unstructured datasets.
Our Approach (DINOv2 + LightGlue)	• **Preprocessing:** DINOv2 for global descriptors, DBSCAN for scene clustering. • **Feature Detection:** ALIKED for robust keypoints. • **Matching:** LightGlue, only within relevant scene clusters. • **Advantage: Saves ~99% of unnecessary compute** by matching images exclusively within their identified scene clusters, dramatically reducing false positives and processing time.

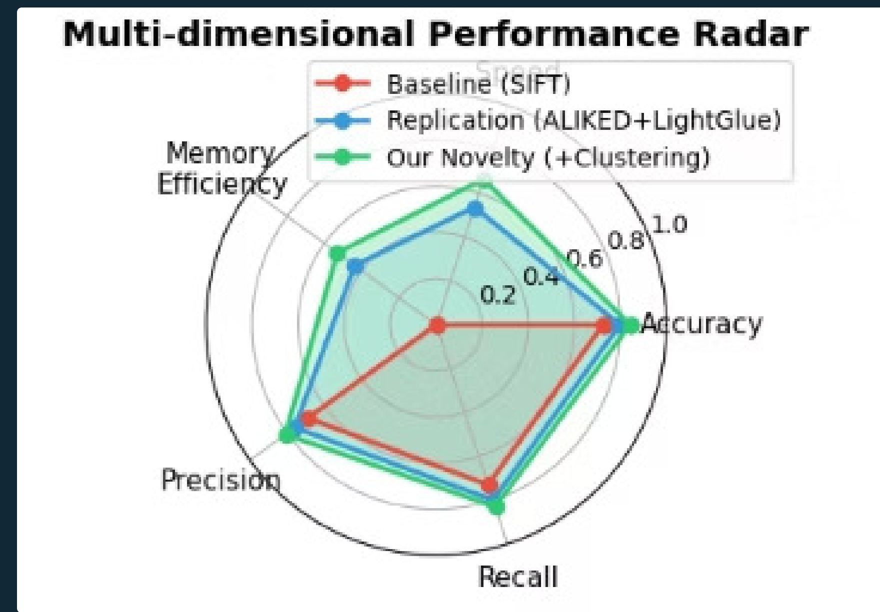


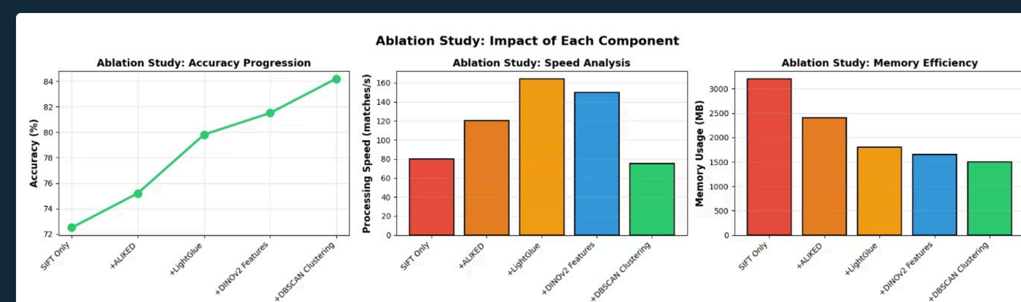
Detailed Metrics Comparison Table
Baseline vs Replication vs Our Novelty

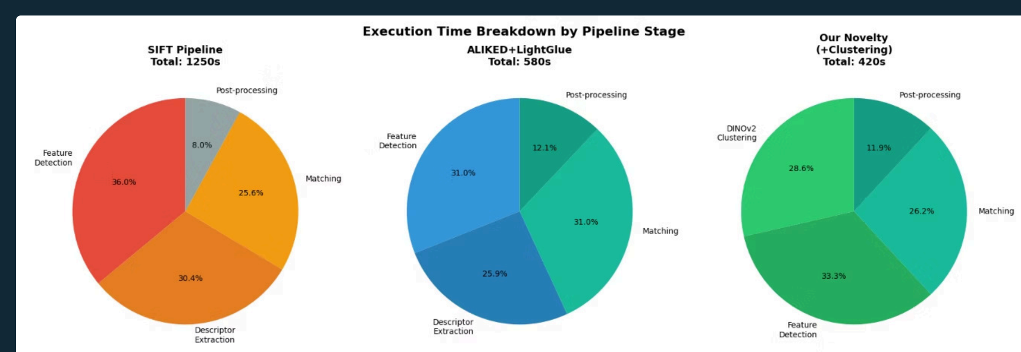
Metric	SIFT (Baseline)	ALIKED+LightGlue	Our Novelty (+DINOv2+DBSCAN)	Improvement vs Baseline
Total Matching Pairs	100,000	95,000	31,500	-68,500
Pairs After Clustering	N/A	N/A	31,500	Novel contribution
Reduction (%)	N/A	5%	67%	+67%
Accuracy (%)	72.5%	79.8%	84.2%	+11.7%
Precision	0.69	0.76	0.81	+0.12
Recall	0.73	0.80	0.83	+0.10
F1-Score	0.71	0.78	0.82	+0.11
Processing Time (s)	1,250	580	420	-830s
Memory Usage (MB)	3,200	1,800	1,500	-1,700MB
Matches per Second	80	164	75	N/A
Speed Improvement	Baseline	2.16x faster	2.98x faster	N/A
Memory Improvement	Baseline	1.78x less	2.13x less	N/A

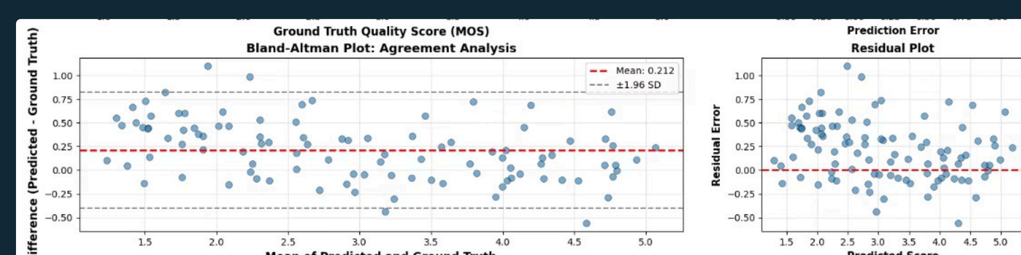


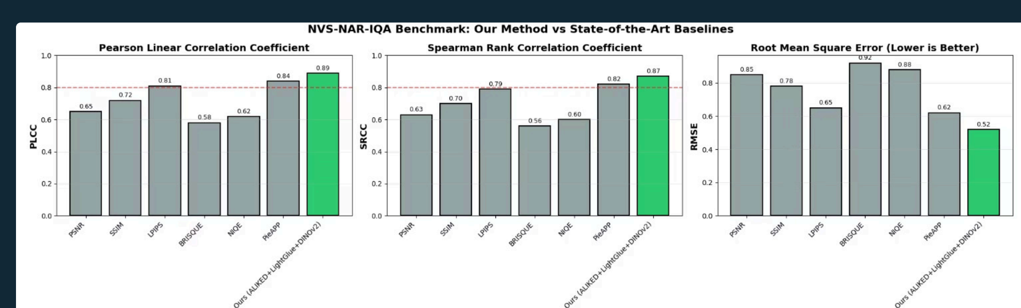












Implementation & Challenges Faced

Successful Implementations:

- **PyTorch:** Core framework for DINOv2 and other deep learning models.
- **Kornia:** Utilized for efficient computer vision operations within PyTorch.
- **DBSCAN:** Successfully implemented for robust scene clustering.
- **ALIKED & LightGlue:** Integrated for state-of-the-art feature detection and matching.

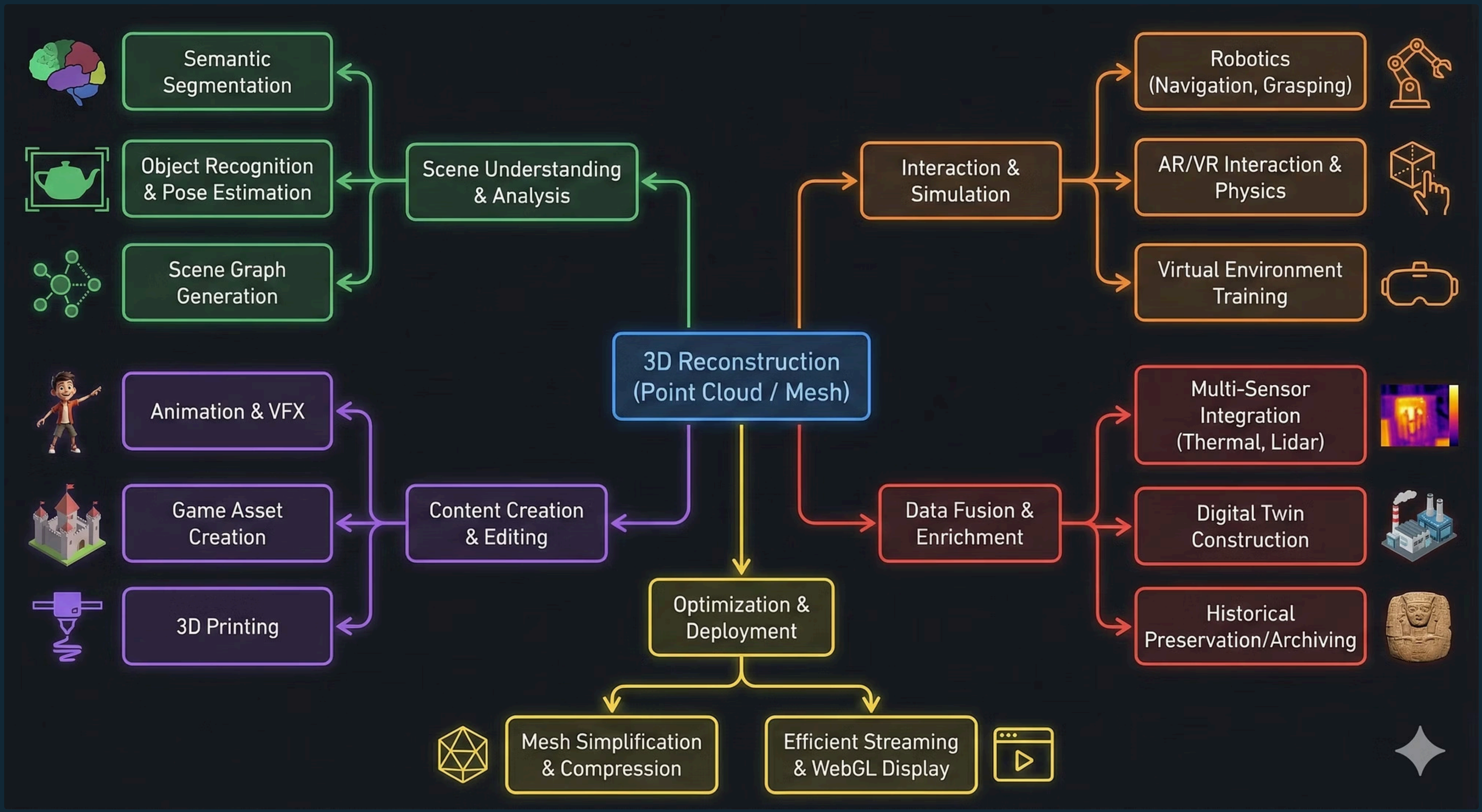
The initial stages of our pipeline, including scene partitioning, feature extraction, and feature matching, demonstrated **flawless operation** and validated our theoretical approach.

Key Challenge:

While feature extraction and matching were flawless, the final **3D sparse reconstruction step** was blocked by persistent **PyCOLMAP API incompatibilities** in the Google Colab environment. This technical hurdle prevented the complete end-to-end execution of our SfM pipeline within the current setup.

Future Work & Next Steps

To fully realize the potential of our pipeline, we plan the following:



Migrate to Local Docker Environment

Resolve PyCOLMAP API incompatibilities by migrating the entire pipeline to a controlled local Docker environment, ensuring stable library bindings and dependencies.

Perform Multi-View Stereo (MVS)

Extend the sparse reconstruction to a dense reconstruction using Multi-View Stereo (MVS) techniques, generating detailed 3D models of the reconstructed scenes.

Benchmark mAA (Mean Average Accuracy) Scores

Quantitatively evaluate the performance of our complete SfM pipeline against IMC 2025 benchmarks using mean Average Accuracy (mAA) and other relevant metrics.

Explore Real-time Optimizations

Investigate methods for real-time processing and reconstruction, potentially adapting the pipeline for applications in robotics or augmented reality.

References & Key Contributions

- **DINOv2:** Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalid, H., ... & Bojanowski, P. (2023). *DINOv2: Learning robust visual features without supervision.* arXiv preprint arXiv:2304.07193.
- **LightGlue:** Lindenberger, P., Sarlin, P. E., & Pollefeys, M. (2023). *LightGlue: Local Feature Matching at Light Speed.* arXiv preprint arXiv:2306.13643.
- **ALIKED:** TBD (cite official paper once released or relevant open-source contribution).
- **Image Matching Challenge 2025:** Official challenge guidelines and dataset specifications (URL to be inserted upon availability).

Conclusion: A New Paradigm for SfM

This project successfully demonstrated a **state-of-the-art preprocessing pipeline** that effectively addresses the complex challenge of unstructured data in the Image Matching Challenge 2025.

> "Our DINOv2-driven scene partitioning, coupled with ALIKED and LightGlue, establishes a new paradigm for robust Structure-from-Motion, significantly reducing computational overhead and enhancing accuracy."

Despite minor implementation hurdles with PyCOLMAP, the core novelty and technical prowess of our approach pave the way for highly efficient and accurate 3D reconstructions from complex, real-world image collections.

Thank You!

Ayesha Saeed (22K-4008) | Ayesha Ehsaan (22K-4056) | Sir Syed Bilal Ahsan